


SAAS

 DEPARTMENT OF STATISTICS
& ACTUARIAL SCIENCE
The University of Hong Kong

Patrick Poon Lecture Series in Actuarial Science

Using optimal transport to mitigate unfair predictions and quantify counterfactual fairness



by **Professor Arthur Charpentier**

Department of Mathematics

University of Quebec at Montreal

April 10, 2026 (Friday)

11:00 a.m. - 12:00 noon

Room 301, Run Run Shaw Building, HKU

Abstract

Many industries are heavily reliant on predictions of risks based on characteristics of potential customers. Although the use of said models is common, researchers have long pointed out that such practices perpetuate discrimination based on sensitive features such as gender or race. Given that such discrimination can often be attributed to historical data biases, an elimination or at least mitigation, is desirable. With the shift from more traditional models to machine-learning based predictions, calls for greater mitigation have grown anew, as simply excluding sensitive variables in the pricing process can be shown to be ineffective.

In the first part of this seminar, we propose to mitigate possible discrimination (related to so call group fairness, related to discrepancies in score distributions) through the use of Wasserstein barycenters instead of simple scaling. To demonstrate the effects and effectiveness of the approach we employ it on real data and discuss its implications.

In the second part, we will focus on another aspect of discrimination usually called counterfactual fairness, where the goal is to quantify a potential discrimination if that person had not been Black or if that person had not been a woman. The standard approach, called *ceteris paribus* (everything remains unchanged) is not sufficient to take into account indirect discrimination, and therefore, we consider a mutates mutants approach based on optimal transport. With multiple features, optimal transport becomes more challenging and we suggest a sequential approach based on probabilistic graphical models.